

# Xenacoelomorpha Survey Reveals That All 11 Animal Homeobox Gene Classes Were Present in the First Bilaterians

Michael Brauchle<sup>1,2,\*</sup>, Adem Bilican<sup>3,†</sup>, Claudia Eyer<sup>3</sup>, Xavier Bailly<sup>4</sup>, Pedro Martínez<sup>5,6</sup>, Peter Ladurner<sup>7</sup>, Rémy Bruggmann<sup>3,\*</sup>, and Simon G. Sprecher<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Institute of Zoology, University of Fribourg, Switzerland

<sup>2</sup>Department of Biology, Institute of Cell Biology, University of Bern, Switzerland

<sup>3</sup>Department of Biology, Interfaculty Bioinformatics Unit, University of Bern, Switzerland

<sup>4</sup>UPMC-CNRS FR2424, Station Biologique de Roscoff, Roscoff, France

<sup>5</sup>Departament de Genètica, Universitat de Barcelona, Catalonia, Spain

<sup>6</sup>Institut Català de Recerca i Estudis Avancats (ICREA), Barcelona, Spain

<sup>7</sup>Institute of Zoology and Center of Molecular Bioscience Innsbruck, University of Innsbruck, Austria

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: E-mails: [simon.sprecher@unifr.ch](mailto:simon.sprecher@unifr.ch); [michael.brauchle@izb.unibe.ch](mailto:michael.brauchle@izb.unibe.ch); [remy.bruggmann@bioinformatics.unibe.ch](mailto:remy.bruggmann@bioinformatics.unibe.ch).

Accepted: August 7, 2018

**Data deposition:** This project has been deposited at NCBI under the accession numbers for lp: SAMN07276911, Sr: SAMN07276888 and Xb: SAMN07276887.

## Abstract

Homeodomain transcription factors are involved in many developmental processes across animals and have been linked to body plan evolution. Detailed classifications of these proteins identified 11 distinct classes of homeodomain proteins in animal genomes, each harboring specific sequence composition and protein domains. Although humans contain the full set of classes, *Drosophila melanogaster* and *Caenorhabditis elegans* each lack one specific class. Furthermore, representative previous analyses in sponges, ctenophores, and cnidarians could not identify several classes in those nonbilaterian metazoan taxa. Consequently, it is currently unknown when certain homeodomain protein classes first evolved during animal evolution. Here, we investigate representatives of the sister group to all remaining bilaterians, the Xenacoelomorpha. We analyzed three acoel, one nemertodermatid, and one *Xenoturbella* transcriptomes and identified their expressed homeodomain protein content. We report the identification of representatives of all 11 classes of animal homeodomain transcription factors in Xenacoelomorpha and we describe and classify their homeobox genes relative to the established animal homeodomain protein families. Our findings suggest that the genome of the last common ancestor of bilateria contained the full set of these gene classes, supporting the subsequent diversification of bilaterians.

**Key words:** homeobox genes, hox, acoel, xenacoelomorpha, evolution, patterning genes.

## Introduction

Xenacoelomorpha unites acoel, nemertodermatid, and *Xenoturbella* species. Most recent phylogenetic analyses place this group of animals at the base of bilaterians (Hejnlol et al. 2009; Cannon et al. 2016; Rouse et al. 2016). Previously, there has been considerable debate about the placement of representatives of this group (Hejnlol et al. 2009; Philippe et al. 2011) but the new molecular data corroborate a scenario that

was already favored by previous studies (Westblad 1949; Hyman 1951; Ruiz-Trillo et al. 1999; Achatz et al. 2013). To gain further insight into the evolutionary history of bilaterians Xenacoelomorpha serve as an important group to investigate, because a subset of their features might represent ancestral character states of bilaterians. Currently, there are no available genome assemblies for any of these species and very few are cultivatable in the laboratory. Here we

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

investigate the homeobox gene content in representatives of the Acoela, Nemertodermatida, and Xenoturbella clades, allowing us to identify a minimal set of homeodomain containing proteins in the genome of the last common bilaterian ancestor.

The homeodomain is an ancient protein domain containing around 60 aa and functions as a DNA-binding domain (McGinnis et al. 1984; Scott and Weiner 1984; Gehring et al. 1994). It can be found in plants, fungi and metazoans and has undergone independent expansions in those groups. Soon after its discovery it was recognized that the homeodomain can be identified in all surveyed animal groups (Gehring et al. 1994) and it is now established that these proteins play key roles in cell fate identity (Zheng and Chalfie 2016). Homeodomain proteins are expressed in practically all animal cells and a subset of them, the Hox genes, plays important roles in the patterning of bilaterian animals, for example, by specifying the position along the central nervous system (L.Z. Holland 2015).

Recent genome sequencing projects have investigated the full set of recognizable homeodomain containing proteins in many organisms. Bacteria and Archaea do not contain homeodomain proteins but a similar helix-turn-helix motif, likely representing the ancestral domain structure (Holland 2013). In fungi and single cell organisms the number tends to be small whereas over 100 homeodomain proteins are found in some flowering plants (Bürglin and Affolter 2016). Around 100 proteins containing this domain are also found in protostomes and deuterostomes, of which vertebrates contain about 250 such genes due to 2 rounds of genome duplication (Holland 2013). Because of the high interest in homeodomain proteins they have been very well annotated in a number of cases. For example, in humans the 255 functional homeodomain proteins are grouped into 104 gene families (Zhong and Holland 2011a). More recently updated homeobox gene annotations are also available from *Drosophila melanogaster* (Bürglin and Affolter 2016) and *Caenorhabditis elegans* (Hench et al. 2015), both containing just over 100 homeodomain proteins. Moreover, in other basal animal clades similar efforts have led to a good understanding of the homeobox gene content, for example in representatives of sponges (Larroux et al. 2008), ctenophores (Ryan et al. 2010), placozoans (Srivastava et al. 2008), or cnidarians (Ryan et al. 2006).

Homeodomain proteins can be grouped into distinct classes according to shared features in addition to the homeodomain sequence itself. Within animals 11 classes of homeodomain proteins have been defined; these include ANTP (with HOXL and NKL subclasses), PRD, LIM, POU, HNF, SINE, TALE, CUT, PROS, zinc finger (ZF), and CERS domain proteins. The naming reveals the additional domains present: For example, the ZF class contains homeobox genes with additional ZF motifs, or the LIM class genes also contain LIM domains and so forth. Some of these

classes still contain many members and generally such classes are then further subdivided into gene families. In line with previous work, we here add genes to a particular family if they descend from a gene that is, for example, still recognized in *D. melanogaster* and human and therefore was already present in early bilaterians (Holland 2013; Bürglin and Affolter 2016).

The class of homeobox genes gathering most attention are the members of the Hox gene cluster, a group of genes shared by bilaterian animals and corresponding to the subset of ANTP genes expressed along the anteroposterior axis during vertebrate development (Holland 2013). These Hox genes have been studied in many animal models and often they have been cloned and analyzed prior to the respective genome sequencing project. For example, the organization of the single amphioxus Hox cluster suggested that it represents the ancestral form of the vertebrate Hox cluster (García-Fernández and Holland 1994). Moreover, studies of planarian Hox genes indicated that an already elaborate set must have existed in early protostomes (Balavoine and Telford 1995). These studies were primarily possible because homeodomain proteins are very well conserved; above all within the homeodomain itself. This feature has been very useful to establish homologies also across groups of distantly related animals and some of the first molecular phylogenies trying to establish deep animal relationships, for example, among protostomes, have been constructed based on homeodomain proteins (de Rosa et al. 1999). Homeodomain sequences are still used to reconstruct animal group relationships, for example among molluscs (Biscotti et al. 2014), to group rotiferans (Frobies and Funch 2017) or are used to test different evolutionary codon models (Matassi et al. 2015). However, it is clear that certain homeobox genes also underwent expansions in particular lineages, for example, the Hox genes in Lepidoptera, TALE genes in molluscs or PRD genes in placental mammals and such expansions have sometimes been linked to evolutionary novelties (Holland et al. 2017).

Here we used next-generation sequencing of total RNA samples to identify and annotate expressed homeobox genes in three acoel and one *Xenoturbella* species together with the analysis of additional published Xenacoelomorpha transcriptomes (Cannon et al. 2016). We go on to compare the homeodomain protein content of these xenacoelomorphs to other animal clades and are able to classify the majority of these genes into known animal homeodomain classes and families. We show that the stem xenacoelomorph must have possessed members of all of the 11 homeodomain classes and that they have been retained and are recognizable in currently living representatives. Our work establishes that the last common ancestor of Bilateria was equipped with the full set of homeodomain classes which can be interpreted as support for the idea of an important role for homeobox genes in bilaterian evolution (P.W. Holland 2015).

## Materials and Methods

### RNA Preparation

*Isodiametra pulchra* worms were cultured according to standard conditions (De Mulder et al. 2009). Briefly, worms were kept in nutrient enriched f/2 artificial sea water and kept in an incubator at 19 °C on a 12 h day/night cycle. Worms of all ages were then collected and RNA was purified using a Qiagen RNA purification kit. For *Symsagittifera roscoffensis* gravid worms were kept and released egg cocoons were collected. Upon hatching of the larvae they were pooled and used for an RNA extraction as above before they took up the symbiotic algae *Tetraselmis convolutae*. *Xenoturbella bocki* RNA was extracted from whole animals freshly collected near the Kristineberg research station.

### Sequence Assembly, Homeodomain and Family Membership Identification

We sequenced the poly-adenylated transcriptome (mRNA) of *I. pulchra*, *S. roscoffensis* and *X. bocki* on an Illumina HiSeq 3000. We generated a total of 84 930 312, 61 857 037, and 116 117 207 paired-end 150 bp long reads for *I. pulchra*, *S. roscoffensis*, and *X. bocki*, respectively, and these data sets have been uploaded to NCBI (accession numbers for Ip: SAMN07276911, Sr: SAMN07276888, and Xb: SAMN07276887). In addition, we downloaded the publicly available data set for *Hofstenia miamia* sampling expressed RNA from mixed stages SAMN02690669 (Srivastava et al. 2014), which represented a total of 87 544 989 paired-end 80 bp long reads as well as the *Nemertoderma westbladi* data set, part of the PRJNA295688 Xenacoelomorpha project (Cannon et al. 2016). Next, we performed a trimming of low quality bases where leading or trailing bases with a quality score lower than three were removed as well as a 4 bp sliding window search for low quality bases where reads were cut if the average quality score dropped below 20. The trimmed reads were then used to perform a de novo transcriptome assembly using Trinity (Grabherr et al. 2011). Open Reading Frames were identified based on the TransDecoder tool and searched for the “Homeobox” term from the PFAM database (Finn et al. 2016). Sequences of all homeodomain containing Open Reading Frame’s (ORF’s) can be found in [supplementary data S6, Supplementary Material](#) online. To avoid false positive homeodomain identification, we applied the following selection criteria: A *P*-value below 0.001 and an initial length of at least 40 amino acids. We then performed a manual curation of identified homeodomain containing proteins and supplemented this list with additional, manually retrieved, homeodomain sequences of partial ORF’s. Next we performed BLAST searches on NCBI and the HomeoDB2 database to identify affiliations to known homeobox gene classes and families. A gene was placed into a particular family if all of the top ten hits of the blast results against HomeoDB2

identified the same family. Genes not fulfilling this criterion are indicated as ambiguous in [supplementary figure S3, Supplementary Material](#) online. In addition, we searched the ORFs for additional PFAM protein domains to support the class identification ([supplementary fig. S3, Supplementary Material](#) online). As described, and in selected cases only, BLAST searches were also conducted for other species to check if other genomes harbor genes of interest. The *I. pulchra* *Hox5* homolog was manually assembled from DN31614 and DN17434, as an intron remained in the original transcript assembly likely due to unfinished intron removal during splicing. Likewise, the *I. pulchra* *Gsx* homolog (DN47321) contained introns that were manually removed.

### Generation of the Phylogenetic Trees

To analyze the phylogenetic relation between the different Xenacoelomorpha homeobox genes, we used the maximum likelihood approach based on subsets of the 60 aa sequences of the recovered homeodomains listed in [supplementary figure S4 and table S11, Supplementary Material](#) online. All the sequences were first aligned using Prank (Loytynoja 2014) and if needed poorly aligned regions were removed using Trimal (Capella-Gutiérrez et al. 2009). The final trees were then computed using PhyML (Guindon et al. 2010) with 500 bootstraps and in some cases rooted as indicated in the figure legends. We constructed trees with acoel sequences alone, acoel together with human sequences (Holland et al. 2007) as well as from five Xenacoelomorpha data sets together with *C. elegans*, *D. melanogaster*, *Branchiostoma floridae*, *C. gigas*, and *Nematostella vectensis* both for ANTP as well as for the rest of the homeodomain classes ([supplementary table S11, Supplementary Material](#) online). All trees show similar within-family groupings of acoel or Xenacoelomorpha sequences, whereas the position of families within a class can change. To independently assess orthology relationships we additionally performed reciprocal BLAST analysis for a subset of genes confirming our classification (data not shown).

## Results

### Homeodomain Protein Content in Xenacoelomorpha Species

To gain insight into the number of homeodomain classes and proteins present early on during bilaterian evolution we ultimately chose to focus on five different Xenacoelomorpha species. As there is currently no assembled genome of any such species available we performed RNA sequencing on *I. pulchra*, *S. roscoffensis*, and *X. bocki*. *Isodiametra pulchra* have been cultivated in the lab for many years (De Mulder et al. 2009), *S. roscoffensis* can be easily collected in large numbers, for example, in Brittany (Bailly et al. 2014) and *X. bocki* can be collected in the Gullmarn Fjord in Sweden (Westblad 1949). In addition, we performed detailed analyses

**Table 1.**

Number of Xenacoelomorph Genes in Each Homeobox Gene Class in Comparison to Other Model Systems

| Gene Class <sup>a</sup> | <i>B. floridae</i> <sup>b</sup> | <i>D. melanogaster</i> <sup>b</sup> | <i>C. elegans</i> <sup>b</sup> | <i>C. gigas</i> <sup>b</sup> | <i>I. pulchra</i> <sup>c</sup> | <i>S. roscoffensis</i> <sup>c</sup> | <i>H. miamia</i> <sup>c</sup> | <i>N. westbladi</i> <sup>c</sup> | <i>X. bocki</i> <sup>c</sup> | <i>N. vectensis</i> <sup>b</sup> |
|-------------------------|---------------------------------|-------------------------------------|--------------------------------|------------------------------|--------------------------------|-------------------------------------|-------------------------------|----------------------------------|------------------------------|----------------------------------|
| ANTP                    | 60                              | 47                                  | 32                             | 51                           | 32                             | 28                                  | 30                            | 31                               | 27                           | 74                               |
| PRD                     | 27                              | 26                                  | 17                             | 30                           | 16                             | 16                                  | 19                            | 13                               | 17                           | 33                               |
| LIM                     | 7                               | 7                                   | 7                              | 8                            | 8                              | 8                                   | 9                             | 13                               | 6                            | 4                                |
| TALE                    | 9                               | 8                                   | 5                              | 23                           | 9                              | 9                                   | 8                             | 9                                | 7                            | 6                                |
| POU                     | 7                               | 5                                   | 3                              | 4                            | 4                              | 3                                   | 4                             | 4                                | 3                            | 5                                |
| SINE                    | 3                               | 3                                   | 4                              | 3                            | 3                              | 3                                   | 3                             | 2                                | 4                            | 5                                |
| CUT                     | 4                               | 3                                   | 8                              | 6                            | 2                              | 2                                   | 1                             | 3                                | 3                            | 1                                |
| ZF                      | 5                               | 2                                   | 2                              | 4                            | 3                              | 3                                   | 3                             | 1                                | 1                            | 0                                |
| HNF                     | 4                               | 0                                   | 1                              | 0                            | 1                              | 1                                   | 1                             | 1                                | 1                            | 1                                |
| CERS                    | 1                               | 1                                   | 0                              | 1                            | 1                              | 1                                   | 1                             | 4                                | 2                            | 1                                |
| PROS                    | 1                               | 1                                   | 1                              | 1                            | 1                              | 1                                   | 1                             | 2                                | 1                            | 0                                |
| Ambiguous               | 3                               | 0                                   | 23                             | 3                            | 3                              | 3                                   | 1                             | 3                                | 2                            | 4                                |
| Total                   | 131                             | 103                                 | 103                            | 134                          | 83                             | 78                                  | 81                            | 86                               | 74                           | 134                              |

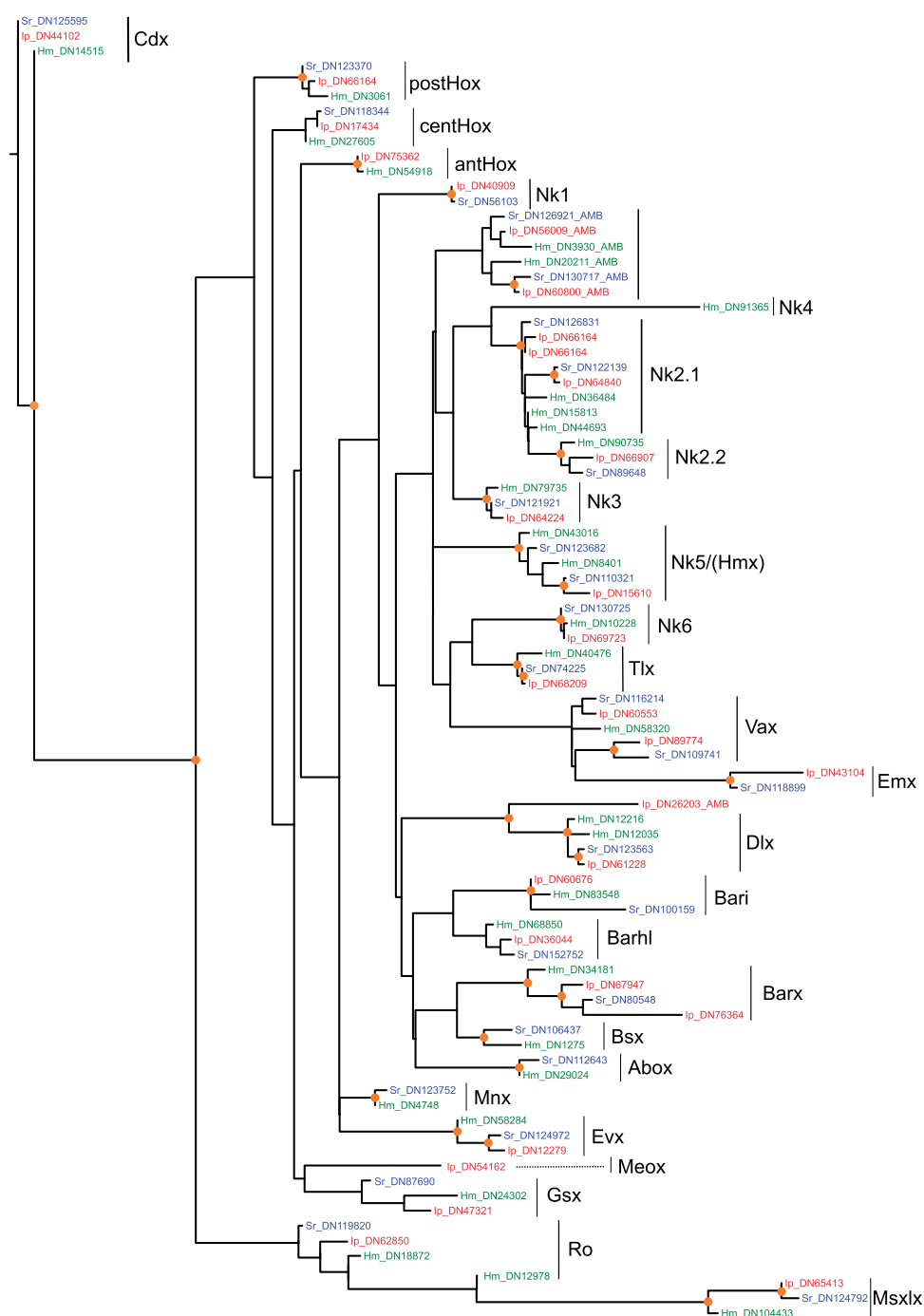
<sup>a</sup>Listed are the 11 animal homeobox gene classes.<sup>b</sup>These data come from whole genome sequencing.<sup>c</sup>These data come from transcriptome analyses.

of the available transcriptomes of a third acoe species *H. miamia* (Srivastava et al. 2014) and of the nemertodermatid *N. westbladi* (Cannon et al. 2016). Upon Illumina sequencing we searched the newly assembled transcriptomes using the PFAM homeodomain sequence to identify the genes containing a homeodomain in the five analyzed species. Overall we identified 83, 78, 81, 86, and 74 high confidence homeobox containing genes in *I. pulchra*, *S. roscoffensis*, *H. miamia*, *N. westbladi*, and *X. bocki*, respectively, (table 1; supplementary table S1, Supplementary Material online). Moreover, we also surveyed additional available Xenacoelomorpha transcriptomes for their homeodomain containing proteins (supplementary table S2, Supplementary Material online) (Cannon et al. 2016).

Next, we set out to classify these newly identified homeodomain proteins into known animal homeodomain classes and families. To this end, we first performed BLAST analyses with every identified Xenacoelomorpha homeodomain sequence at NCBI and Homeodb2 to reveal likely class and family memberships relative to well annotated genomes (supplementary figs. S3 and S4, Supplementary Material online). In addition, we performed reciprocal BLAST analyses that confirmed the results from such searches (data not shown). As done in previous studies, we then grouped the sequences in ANTP and non-ANTP homeoboxes, aligned them, and initially generated phylogenetic trees of only the acoe sequences alone (figs. 1 and 2, see Material and Methods). Of the 61 families predicted by the BLAST analysis for acoels, 43 are supported by a bootstrap value over 70%, confirming the close relatedness of individual family members (orange nodes in figs. 1 and 2). Furthermore, these analyses identified only 3 families (Meox, Nk4 and Pou6) that we found in only a single acoe species while 4 families contained members of 2 species (antHox, Nk1, Emx, and Shox) and 54 families have at least

one member of all three acoe species, suggesting a good coverage of the expressed homeodomain containing sequences in our acoe data set (figs. 1 and 2).

Next we assembled and aligned the homeodomain sequences from five Xenacoelomorpha species together with other species that have well annotated sets of homeodomain protein sequences (supplementary fig. S4, Supplementary Material online). These include representatives of chordates, *B. floridae* (Takatori et al. 2008) and humans (Holland et al. 2007), a lophotrochozoan, *Crassostrea gigas* (Paps et al. 2015), two ecdysozoans, *C. elegans* (Hench et al. 2015) and *D. melanogaster* (Bürglin and Affolter 2016) and a cnidarian, *N. vectensis* (Ryan et al. 2006). We then generated two more sets of phylogenetic trees: One including the three acoels and all human sequences (supplementary figs. S5 and S6, Supplementary Material online) and one of all five Xenacoelomorpha species together with *B. floridae*, *C. gigas*, *C. elegans*, *D. melanogaster* and *N. vectensis* (supplementary figs. S7 and S8, Supplementary Material online). To generate the human tree we included a few additional homeodomain sequences of the particular families lacking in humans. To the ANTP tree we added *Amphioxus Ro*, *Abox*, and *Bari* and to the non-ANTP tree we added *Amphioxus Repo* because humans lack the respective homologs (supplementary figs. S5 and S6, Supplementary Material online). As expected, the ten non-ANTP homeodomain protein classes separate reasonably well into the distinct classes in both sets of trees (supplementary figs. S6 and S8, Supplementary Material online). Despite the fact that there were genome duplications on the lineage leading to humans essentially the same results are obtained with acoe and human sequences alone as with the full data set (supplementary figs. S5–S8, Supplementary Material online). A representative subregion of the full non-ANTP tree illustrates recovered family groupings of metazoan

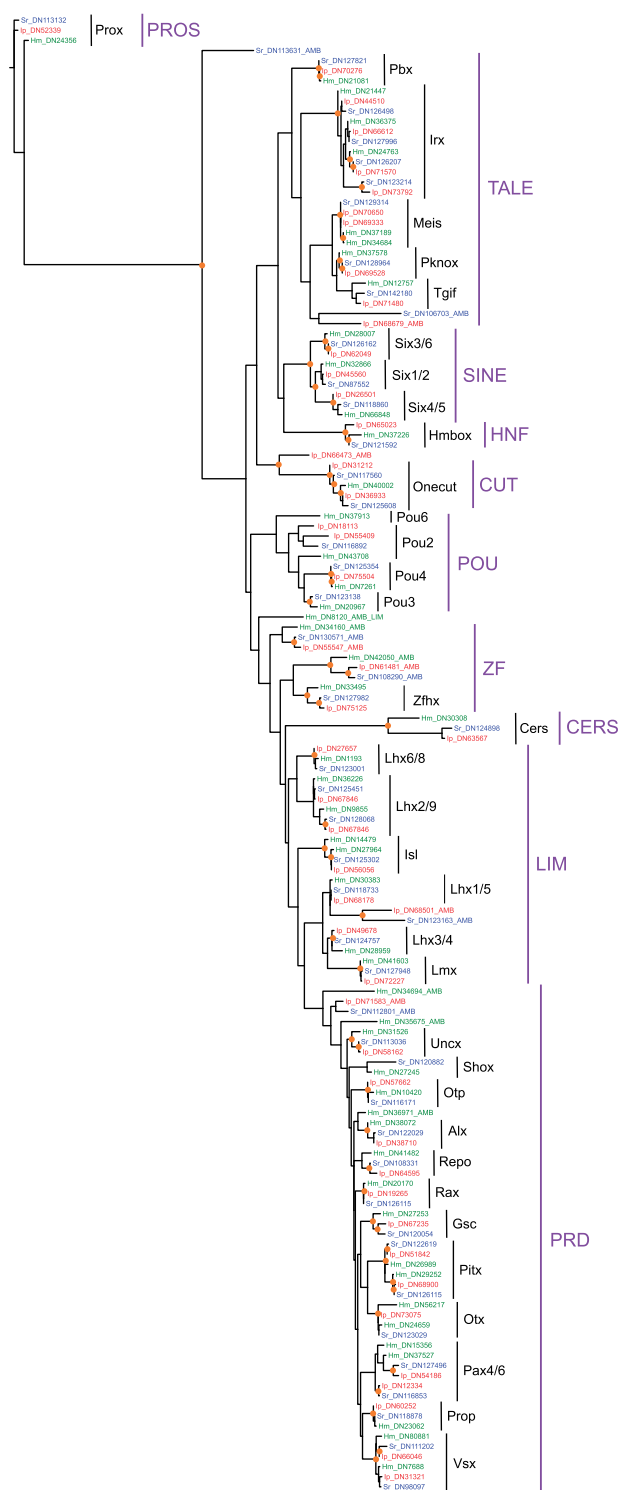


**FIG. 1.**—Sequence relationship among identified ANTP homeodomain sequences of three acoel species. A phylogenetic tree based on the maximum likelihood method, indicating the relationship of all identified acoel ANTP homeodomain proteins and their family classification. Genes extended by \_AMB indicates ambiguous family relationship. Orange labeled nodes have a bootstrap value >70%. Acoel species are color-coded in the following way: Red, *I. pulchra* (Ip); blue, *S. roscoffensis* (Sr); and green, *H. miamia* (Hm). Full sequences are listed in [supplementary fig. S3](#), [Supplementary Material](#) online, alignments are presented in [supplementary fig. S4](#), [Supplementary Material](#) online and trees of acoel sequences together with human or 7 additional animal homeobox data sets in [supplementary figs. S5 and S7](#), [Supplementary Material](#) online, respectively.

homeodomain sequences of the ten analyzed species ([fig. 3](#); [supplementary fig. S8](#), [Supplementary Material](#) online). As expected individual families evolve independently, for

example the Gsc family has a single member in all ten species, *C. elegans* lacks an Otp family member, a Prop family member has not been identified yet in *N. westbladi* and *N. vectensis*,





**FIG. 2.**—Sequence relationship among all non-ANTP homeodomain sequences of three acoel species. A phylogenetic tree, presented as in figure 1, for all non-ANTP homeodomain proteins identified from the three acoel transcriptomes. Full sequences are listed in [supplementary fig. S3, Supplementary Material](#) online, alignments are presented in [supplementary fig. S4, Supplementary Material](#) online and trees of acoel sequences together with human or seven additional animal homeobox data sets in [supplementary figs. S6 and S8, Supplementary Material](#) online, respectively.

whereas the Vsx family has apparent duplications in *C. gigas*, *D. melanogaster*, and *C. elegans* (fig. 3). Altogether these analyses allowed us to classify most Xenacoelomorpha genes first into 1 of the 11 animal homeodomain classes ([table 1](#)) and then to add them to described families therein ([supplementary table S1, Supplementary Material](#) online). Only few genes could not be classified suggesting that these cases may be lineage specific, members of so far unrecognized conserved gene families or belong to rapidly diverging gene families.

To corroborate these initial findings and to independently assess if homeobox genes ended up in the correct class, we searched the identified full-length protein sequences for additional domains apart of the homeodomain itself. As mentioned these features are an important part of homeobox gene classification (Holland 2013; Bürglin and Affolter 2016). We found several such domains, including LIM, POU, ZF, CUT, CERS (TLC; TRAM\_LAG1\_CLN8), and HNF features, validating our initial classification ([supplementary fig. S3, Supplementary Material](#) online).

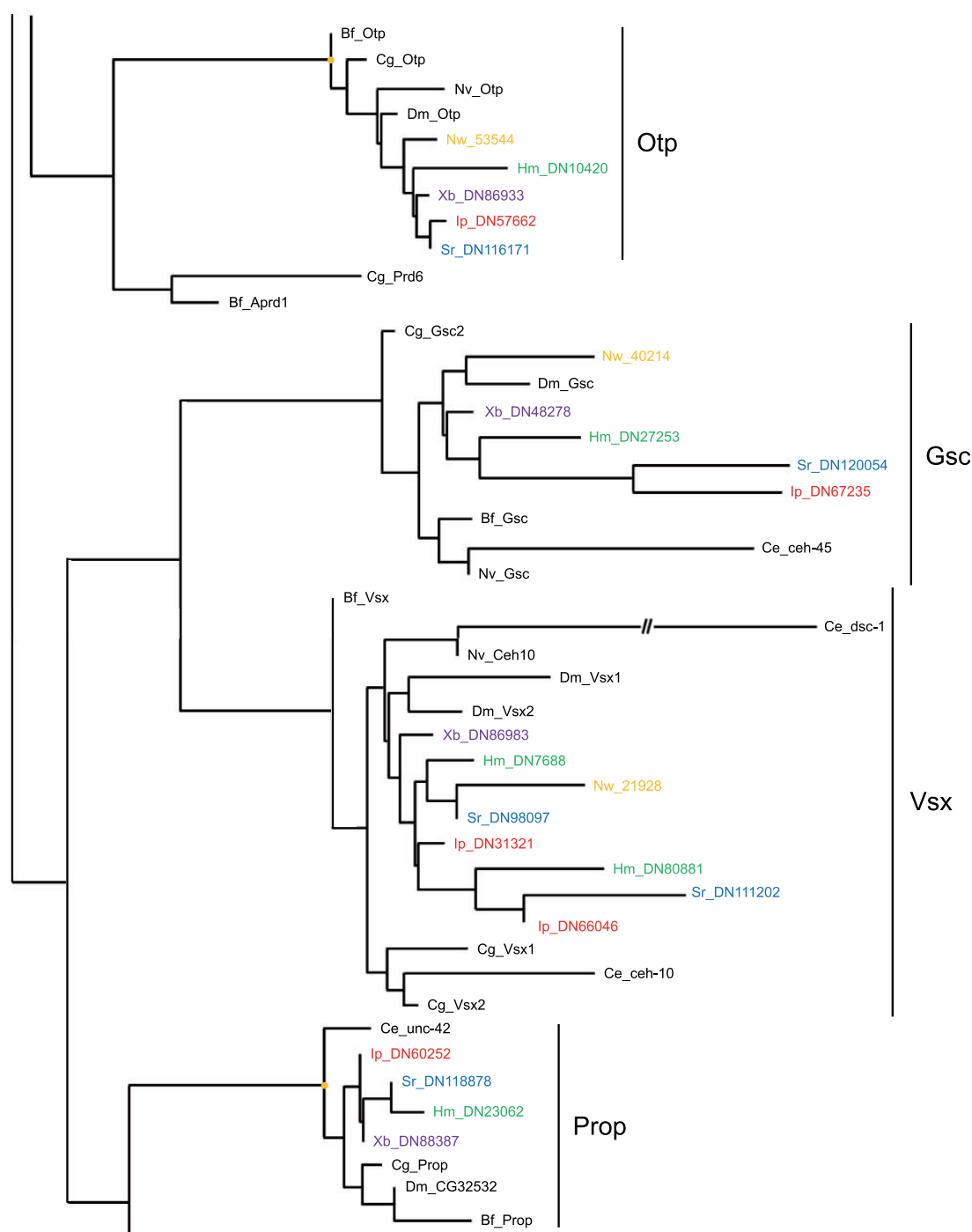
In summary, we identified Xenacoelomorpha members of 68 previously described animal homeodomain families ([supplementary table S1](#) and figs. S4, S7, and S8, [Supplementary Material](#) online) while 13% of genes (53 out of the 402 genes) were not grouped into previously described families. The following sections provide an overview of our classification of the identified genes.

## ANTP Class

ANTP class genes are the largest class of homeobox genes and have so far only been found in animal genomes (Holland 2013). For example, humans contain 100 members in this class alone (Holland et al. 2007). ANTP class genes are generally divided into two subclasses, the HOXL, HOX-like, and NK, NK-like, subclass. Here we followed this subdivision and discuss these subclasses separately.

## HOXL Subclass

HOXL subclass genes have often been implicated in the diversification of animal body plans and therefore they have been heavily investigated in diverse groups of animals (Gellon and McGinnis 1998). Humans contain 14 HOXL families while, for example, *C. elegans* only contains 8 of them (Hench et al. 2015). Previous Hox gene analyses in acoels have recovered three genes, albeit not in a cluster, representing one anterior, *antHox* (previously also called: *Hox1*), one central, *centHox* (*Hox4/5* or *Hox5*), and one posterior, *postHox* (*HoxP*), as well as a single ParaHox gene homologous to *caudal/Cdx* (Cook et al. 2004; Hejnal and Martindale 2009; Moreno et al. 2009). Analyses in *X. bocki* identified an additional central Hox gene member (HoxM2) (Fritzsche et al. 2008; Thomas-Chollier and Martinez 2016). Our transcriptome analysis identified these four genes in the three analyzed acoel



**FIG. 3.**—Xenacoelomorpha homeobox sequences group with described animal homeodomain families. A subregion of a phylogenetic tree (presented as in fig. 1, complete tree [supplementary fig. S8, Supplementary Material](#) online) showing family relationships of four PRD families among five Xenacoelomorpha (three acoels [Ip, Sr, and Hm] with *X. bocki* [Xb, purple] and *N. westbladi* [Nw, orange]), and four additional bilaterian species (*B. floridae* [Bf], *C. gigas* [Cg], *D. melanogaster* [Dm], *C. elegans* [Ce]) and *N. vectensis* (Nv).

species with the exception of the *antHox* from *S. roscoffensis* (also see Discussion), as well as a putative additional central *X. bocki* Hox gene. Moreover, we were able to identify a previously unrecognized homolog of another ParaHox gene,

*Gsx*, in all three acoel species and *Xenoturbella*. Despite a previous report we did not identify central *N. westbladi* homologs, but report an additional putative postHox genes in this species (Jimenez-Guri et al. 2006). In addition, clear *Evx*,

Gbx, Meox, Mnx, and Ro family members were identified, bringing the new total of described HOXL families in Xenacoelomorpha to 10 (fig. 1; [supplementary table S1](#) and figs. S4, S5, and S7, [Supplementary Material](#) online). Taken together, we identified the previously not recovered ParaHox gene *Gsx* and describe several novel HOXL subclass family members in Xenacoelomorpha.

### NKL Subclass

Here we group the identified Xenacoelomorpha NKL subclass genes into any of the described 31 NKL subclass families; the 23 families from humans (Holland et al. 2007) and 8 additional families from other well annotated animal genomes that are not found in humans (Zhong and Holland 2011b).

Among the families found in humans we identified homologs in Xenacoelomorpha for Barhl, Barx, Bsx, Dbx, Dlx, Emx, Hhex, Msx, Nk1, Nk2.1, Nk2.2, Nk3, Nk4, Nk5, Nk6, Nk7, Tlx, and Vax (fig. 1; [supplementary table S1](#) and figs. S4, S5, and S7, [Supplementary Material](#) online). Among the families not found in humans we identified homologs in Xenacoelomorpha for the Abox, Bari, and Msxlx families (fig. 1; [supplementary table S1](#) and figs. S4, S5, and S7, [Supplementary Material](#) online). For a few NKL subclass members we could not unambiguously identify the family relationship. This was not unexpected as, for example, also in *N. vectensis* only 25 of the 54 identified NKL subclass genes could be placed into a human family; although at that time only 17 human classes were recognized (Ryan et al. 2006). All together we found members of 21 previously described families, similar to the 17 present in *N. vectensis* (Ryan et al. 2006) and 20 in *C. elegans* (Hench et al. 2015) but less than the 26 identified in *D. melanogaster* (Bürglin and Affolter 2016).

### PRD Class

Many PRD class genes encode a PAIRED domain (Bopp et al. 1986) while some additional homeobox genes without the PAIRED domain (sometimes called PRD-like) also group with the PRD class phylogenetically and are therefore included in this class (Holland et al. 2007). Some proteins containing a PRD domain, like *Pax1/9* orthologs, do not contain a homeodomain and are therefore not included in this analysis. PRD class genes generally contain either a Q, K or S at position 50 in their homeodomain (Galliot et al. 1999) and we identify members of all of these subclasses in our data set.

The PRD class is the second largest class of homeodomain proteins with 31 human families and another 4 families not found in humans (Zhong and Holland 2011b). Our search of Xenacoelomorpha homeodomain sequences recovered putative members of 14 specific PRD families. They include members of the Alx, Drgx, Gsc, Hbn, Otp, Otx, Pax4/6, Pitx, Prop, Repo, Rax, Shox, Uncx, and Vsx families (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#)

online). In addition, a few proteins cannot easily be associated with a previously described family.

This number is significantly lower when compared with the human content of PRD protein families but is similar to the number in *C. elegans*, which contains 12 PRD families (Hench et al. 2015), or *N. vectensis*, which contains 15 PRD families (Ryan et al. 2006). Nevertheless, PRD homeobox genes do represent the second largest class of homeodomain containing genes in Xenacoelomorpha with around 20 members ([table 1](#)).

### LIM Class

The LIM class members are characterized by additional LIM domains, containing tandemly repeated special ZF motifs, usually located upstream of the homeodomain (Hobert and Westphal 2000) and are named after its presence was discovered together in *lin-11*, *isl-1*, and *mec-3* (Freyd et al. 1990). Six LIM families are known from humans (Holland et al. 2007) and here we recovered members of all of these families in all five species, namely *Isl*, *Lhx1/5*, *Lhx2/9*, *Lhx3/4*, *Lhx6/8*, and *Lmx* except for *Lhx6/8* in *N. westbladi* (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). In addition, putative paralogs were identified in some cases, as well as an additional protein that does not belong to described LIM families ([supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). *Nematostella vectensis* contains four LIM families (Ryan et al. 2006) while sponges contain only three (Larroux et al. 2008). Ctenophores and a placozoan each contain four of the families, although not the same four (Srivastava et al. 2008; Ryan et al. 2010). Together, these analyses suggest that different LIM families were lost in different distantly related animal groups. In addition our data suggest that some LIM families were specifically multiplied in some Xenacoelomorpha species like *N. westbladi* or *D. longitubus* ([supplementary tables S1 and S2](#), [Supplementary Material](#) online).

### POU Class

The POU class members contain an approximately 75 amino acid DNA-binding domain upstream of the homeodomain (Herr et al. 1988). There are six human classes (Ryan and Rosenfeld 1997) and an additional highly divergent family without POU domain, *Hdx*, that weakly groups with the POU class (Holland et al. 2007). A previous analysis in an acoel, *Neochildia fusca*, has already reported Pou3 and -4 homologs (Ramachandra et al. 2002). Here we can confirm this initial finding, as well as establish that Xenacoelomorpha in addition contain members of the Pou2 and -6 families (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). This is similar to *N. vectensis* (Ryan et al. 2006), whereas representatives of sponges have three and placozoans and ctenophores only two of the



human families; again suggesting differential loss of certain families in different ancient animal groups.

### HNF Class

HNF class genes have an atypical and extended homeodomain sequence and were identified with the characterization of a liver specific transcription factor, *Hnf1*, from rats (Chouard et al. 1990). Two gene families containing 3 genes can be found in humans (Holland et al. 2007). We identified one HNF class gene each, although the ends of the atypical homeodomain, containing around 15–20 extra aa between the 2nd and 3rd alpha helices, are difficult to define (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). No HNF class gene is found in *D. melanogaster* (Bürglin and Affolter 2016), whereas *C. elegans* also has a single HNF class gene (Hench et al. 2015). An HNF gene is present in *N. vectensis* (Ryan et al. 2006) and Placozoa (Srivastava et al. 2008) but not in ctenophores (Ryan et al. 2010). The HNF was initially not identified in *Amphimedon queenslandica* (Larroux et al. 2008) although a current BLAST search suggests that a HNF gene is present (XP\_003390473.1; [supplementary fig. S9](#), [Supplementary Material](#) online). These results indicate that the origin of an HNF class gene predates bilateria.

### SINE Class

The SINE class homeodomain proteins contain a conserved SIX domain which is 115 aa in length and are named after the *D. melanogaster* gene *sine oculis*, one of the three SINE class genes in *D. melanogaster* (Seo et al. 1999). In addition, they contain a characteristic lysine at position 50. These three gene families can also be recognized in humans (Holland et al. 2007) and we here establish that all three are also found in Xenacoelomorpha (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). *Nematostella vectensis* likewise contains members of all three human families (Ryan et al. 2006) and so do ctenophores (Ryan et al. 2010). Sponges have only one family (Larroux et al. 2008) and Placozoa two (Srivastava et al. 2008) indicating that different SINE families were already present early on during animal evolution.

### TALE Class

Homeodomains of the TALE (three amino acid loop extension) class contain, as the name indicates, three additional amino acids between helix 1 and 2 of the homeodomain resulting in a 63 aa homeodomain (Bertolino et al. 1995). Such homeodomains can also be found in single cell eukaryotes and plants and represent the ancient homeodomain architecture and have also been called a superclass. In our analysis we treat them as a class and follow the previously suggested subdivision of its members into six human families (Holland et al.

2007). Here we identify members of five of the six human classes in Xenacoelomorpha, that is *Irx*, *Meis*, *Pbx*, *Pknox*, and *Tgif* with the exception being the *Mkx* family (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). This family is also not found in *C. elegans* (Hench et al. 2015) or *N. vectensis* (Ryan et al. 2006) but it is present in *D. melanogaster* (Bürglin and Affolter 2016).

### CUT Class

CUT class homeodomain proteins contain a CUT domain, another DNA-binding domain, upstream of the homeodomain and thereby define this class (Bürglin and Cassata 2002). Three human families have been described (Holland et al. 2007) while a different set of families can be found in *D. melanogaster* (Bürglin and Affolter 2016) or amphioxus (Holland et al. 2008). In our Xenacoelomorpha transcriptomes we find one CUT class gene family homologous to the human *Onecut* homeodomain family (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). A putative *Onecut* homolog is also found in a cnidarian (Nv XP\_001637067.1) (Ryan et al. 2006), but not in studied members of poriferans (Larroux et al. 2008), placozoans (Srivastava et al. 2008), or ctenophores (Ryan et al. 2010). This indicates that CUT class genes first appeared in the cnidarian-bilaterian ancestor.

### CERS Class

CERS class proteins function in ceramide synthesis and are predicted to be transmembrane proteins located in the ER. Only five, (CERS2-6), of the six human CERS proteins identified contain a homeodomain that is highly divergent (Holland et al. 2007) and that can be almost completely deleted without affecting their function (Mesika et al. 2007); consequently, it is currently unclear if they play a role during transcription. There are no CERS class genes that also contain a homeodomain in *C. elegans* (Hench et al. 2015). All our Xenacoelomorpha transcriptomes contain at least one CERS class family member containing a homeodomain (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). No CERS homolog has been reported in *A. queenslandica* (Larroux et al. 2008), although current *A. queenslandica* BLAST searches identify a possible homolog with a truncated homeodomain (ceramide synthase 1-like [XP\_003382832.1]; [supplementary fig. S10](#), [Supplementary Material](#) online). Similarly a *Cers* homolog was not reported in the cnidarian homeodomain analysis (Ryan et al. 2006), but a current BLAST search identifies *N. vectensis* XP\_001635937.1, although with a poorly aligned homeodomain. Currently analyzed ctenophores and placozoans do not seem to contain a CERS homolog (Srivastava et al. 2008; Ryan et al. 2010). Nevertheless, a more detailed analysis has to be conducted to better understand the evolutionary history of CERS class.

### PROS Class

PROS homeobox genes contain a homeodomain extended by 3 aa, although in a different position to the insertion in TALE homeodomains, in addition to the PROS domain that is found c-terminally and was originally identified in the *D. melanogaster Prospero* gene (Chu-Lagraff et al. 1991). We identify *Prospero* homologs in all Xenacoelomorpha (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). *Prospero* has not been found in representatives of sponges (Larroux et al. 2008), cnidarians (Ryan et al. 2006), ctenophores (Ryan et al. 2010), or placozoans (Srivastava et al. 2008). This suggests that *Prospero* is a bilaterian specific homeobox gene.

### Zinc Finger Class

Zinc finger class homeodomains contain C<sub>2</sub>H<sub>2</sub> or C<sub>2</sub>H<sub>2</sub>-like ZFs, typically also involved in DNA binding, in addition to the homeodomain. We identified a single large ZF class protein in all five species containing four homeodomains most similar to the human Zfhx3/4 family that also contains four homeodomains (fig. 2; [supplementary table S1](#) and figs. S4, S6, and S8, [Supplementary Material](#) online). In addition, there are a couple of other proteins with similarity to ZF-type homeodomains in our transcriptomes that we cannot easily group into previously described families. No clear homologs of ZF class proteins are described in poriferans (Larroux et al. 2008), cnidarian (Ryan et al. 2006), or ctenophores (Ryan et al. 2010), nor are they currently found by BLAST searches in those phyla, suggesting that also this class evolved at the base of bilaterians.

## Discussion

### Around 80 Identified Homeodomain Proteins in Five Xenacoelomorpha Transcriptomes

Previously identified Hox and ParaHox genes in acoels include *antHox*, *centHox*, *postHox*, and *Cdx* (Cook et al. 2004; Hejnal and Martindale 2009; Moreno et al. 2009) while four Hox genes have been reported from *X. bocki* (Fritzsch et al. 2008; Thomas-Chollier and Martinez 2016). We recovered 11 out of these 12 possible hits (=92%) in the three acoel species investigated, and all 4 previously reported Hox genes from *X. bocki*, giving us the confidence to analyze the data set in more detail. It will have to be confirmed that the additional central Hox gene we report here from *X. bocki*, as well as the two *postHox* from *N. westbladi*, are real; previous analyses had contamination issues and we cannot formally exclude similar problems (Thomas-Chollier and Martinez 2016). Despite a previous report we did not identify central *N. westbladi* genes, certainly more sampling should be performed to properly assess the previous findings (Jimenez-Guri et al. 2006). Importantly though, we identified all 11 homeobox gene classes in all 5 analyzed transcriptomes and most

of the identified gene families (63 out of 68) are found in more than one of those species ([supplementary table S1](#), [Supplementary Material](#) online). We are therefore confident that our transcriptomes represent a good coverage of expressed homeodomain proteins. Some of the genes we find do not fall into monophyletic groups with members of previously established homeobox gene families. This likely indicates novel or significantly diverged gene families, something that has also been observed in other analyzed species. Moreover, and only with full genome data at hand, will we be able to assess the total number of homeobox genes in specific Xenacoelomorpha species.

### Xenacoelomorpha Homeobox Gene Content from an Evolutionary Perspective

One interesting observation is that more homeobox genes were identified in *N. vectensis*, now 134 including the putative ONECUT and CERS homeodomain proteins (although not from all the classes) than we found in Xenacoelomorpha. For example, more than twice as many genes were found in the NKL subclass in *N. vectensis* than in the acoel species, which is also many more in comparison to *D. melanogaster* and *C. elegans*. This may be due to a rapid expansion of some homeobox gene families in cnidarians similar to the expansion of other important gene families such as *wnt* (Kusserow et al. 2005). We also observe small expansions in our data set, for example in the *Lmx* or *Cers* gene families of *N. westbladi* ([supplementary table S1](#), [Supplementary Material](#) online). The evolutionary reasons for expansions or losses of homeobox genes in specific animal groups are currently not well understood.

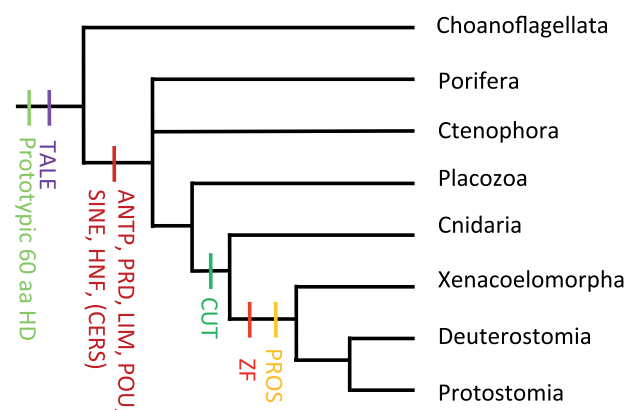
Some homeodomain proteins are lacking from the currently analyzed Xenacoelomorpha. It is tempting to speculate whether these genes were never present, have been lost or are simply not detected. For example, among the ANTP HOXL families we currently find only a single *antHox* member in Xenacoelomorpha that is most similar to *Hox1*. This is in contrast to the situation in *N. vectensis* where several additional *antHox* homologs exist (Chourrout et al. 2006; Ryan et al. 2006). The family relationship of cnidarian *antHox* genes is debated but their existence is used to imply a ProtoHox cluster with more than one anterior Hox gene (Chourrout et al. 2006). Our data would then support a scenario where the *Hox3* homolog was lost in Xenacoelomorpha (Moreno and Martinez 2010).

The overall small number of Hox genes in acoels could indicate a more general trend toward gene loss of derived species within Xenacoelomorpha or could also be due to a fast rate of molecular evolution in certain lineages (Gavilán et al. 2016). In support of this *X. bocki* has two more central *Hox* genes than our analyses in acoels suggesting that the two families were likely lost in acoels, as well as in nemertodermatids. Moreover, our analysis suggests that the

following gene families, Gbx, Hhex, Msx, Nk7, Drgx, and Hbn were lost in acoels relative to *X. bocki*, while Mnx, ro, Abox, Bari, Bsx, Msxlx, Nk3, Vax, and Pou4 are present in acoels but not yet identified in *X. bocki* (supplementary table S1, Supplementary Material online). Ultimately, deeper investigations into several whole genome sequences of basal and derived as well as slow or fast evolving Xenacoelomorpha species are required to better understand the evolution of specific gene families.

### All 11 Homeodomain Classes are Present in Xenacoelomorpha

During evolution a trend toward homeodomain protein amplification has been observed thereby providing opportunities for natural selection while homeobox gene loss has sometimes been linked to morphological simplifications (Holland 2013). The model choanoflagellate *Monosiga brevicollis* contains only two TALE homeodomain transcription factors (King et al. 2008). More recently, it was discovered that some other unicellular eukaryotes possess a few more homeodomain proteins of which some contain PRD type features, or are paired with LIM domains (Grau-Bové et al. 2017); however, these proteins could not be assigned to distinct animal homeodomain families and their evolutionary history deserves further clarifications. The *Amphimedon* genome contains ANTP (NK), PRD, POU, LIM, SINE, and TALE homeodomain classes (Larroux et al. 2008), and here we report a newly identified HNF member in sponges. Currently, it is generally believed that the classical Hox cluster genes, part of ANTP HOXL, evolved only after sponges diverged from eumetazoans (Larroux et al. 2007), but there are studies that suggest that most HOXL genes, maybe with the exception of the putative *Cdx* paralog, have been lost in sponges (Fortunato et al. 2014). Further sequencing of additional sponges might be able to distinguish between these scenarios. A detailed analysis of ctenophore homeodomain proteins suggests that this group lacks members of CUT, PROS, HNF, ZF, and CERS classes (Ryan et al. 2010). Cnidarians are reported to lack members of PROS, ZF, and CERS classes (Ryan et al. 2006) although a more detailed analysis of CERS family members is needed to understand its evolutionary history. Taken together, our results lead us to suggest that Xenacoelomorpha is the oldest animal group of which current descendants still contain all homeodomain transcription factors classes (fig. 4). Why this group of animals would retain all classes is currently unknown especially in the light that other well-studied model systems such as *D. melanogaster* and *C. elegans* lost a class altogether. *Drosophila melanogaster* does not possess any member of the HNF class (Bürglin and Affolter 2016). *Caenorhabditis elegans* lacks members of the CERS homeodomain class (Hench et al. 2015). Our analysis does however establish that Xenacoelomorpha undoubtedly contain all 11 animal homeodomain classes and therefore suggests that the genome of



**FIG. 4.**—Hypothesized appearance of homeobox gene classes during animal evolution. Shown is a simplified phylogenetic tree containing animal groups of interest according to recent molecular data (Cannon et al. 2016; Rouse et al. 2016). The placement of ctenophores in relation to poriferans is not resolved on this tree. Indicated are the points at which homeodomain classes are first identified. Observed losses of certain homeodomain protein classes in specific lineages are not indicated here.

the last common ancestor of Bilateria also contained the full set of these gene classes. This possession of all animal homeodomain classes could have been an important prerequisite for the following radiation of bilaterians. Indeed our data set is in agreement with a previous proposition suggesting that homeobox gene duplications could have played an important role in the Cambrian explosion (P.W. Holland 2015) because we are able to show that the full set of homeobox genes was indeed present early on during the evolution of bilateral animals.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

We would like to thank members of the Sprecher and Bruggmann laboratories, colleagues at the marine biology station in Roscoff and the Sven Lovén Center in Kristineberg for constructive and fruitful discussions. Moreover, J. Cannon and B. Vellutini shared previous data sets with us. A.B. would like to thank Walid Gharib and Romain Studer for fruitful discussions regarding phylogenetic analyses. This work was supported by the Swiss National Science Foundation (51RT-0\_145733) and the University of Fribourg to S.G.S.

### Literature Cited

- Achatz JG, Chiodin M, Salvenmoser W, Tyler S, Martinez P. 2013. The Acoela: on their kind and kinships, especially with nemertodermatids and xenoturbellids (Bilateria incertae sedis). *Org Divers Evol*. 13(2):267–286.

- Bailly X, et al. 2014. The chimerical and multifaceted marine acoel *Symsagittifera roscoffensis*: from photosymbiosis to brain regeneration. *Front Microbiol.* 5:498.
- Balavoine G, Telford MJ. 1995. Identification of planarian homeobox sequences indicates the antiquity of most Hox/homeotic gene sub-classes. *Proc Natl Acad Sci U S A.* 92(16):7227–7231.
- Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG. 1995. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem.* 270(52):31178–31188.
- Biscotti MA, Canapa A, Forconi M, Barucca M. 2014. Hox and ParaHox genes: a review on molluscs. *Genesis* 52(12):935–945.
- Bopp D, Burri M, Baumgartner S, Frigerio G, Noll M. 1986. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of *Drosophila*. *Cell* 47(6):1033–1040.
- Bürglin TR, Affolter M. 2016. Homeodomain proteins: an update. *Chromosoma* 125(3):497–521.
- Bürglin TR, Cassata G. 2002. Loss and gain of domains during evolution of cut superclass homeobox genes. *Int J Dev Biol.* 46(1):115–123.
- Cannon JT, et al. 2016. Xenacoelomorpha is the sister group to Nephrozoa. *Nature* 530(7588):89–93.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Chouard T, et al. 1990. A distal dimerization domain is essential for DNA-binding by the atypical HNF1 homeodomain. *Nucleic Acids Res.* 18(19):5853–5863.
- Chourrout D, et al. 2006. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442(7103):684–687.
- Chu-Lagraff Q, Wright DM, McNeil LK, Doe CQ. 1991. The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in *Drosophila*. *Dev Suppl.* 7(7):205–285.
- Cook CE, Jimenez E, Akam M, Salo E. 2004. The Hox gene complement of acoel flatworms, a basal bilaterian clade. *Evol Dev.* 6(3):154–163.
- De Mulder K, et al. 2009. Characterization of the stem cell system of the acoel *Isodiametra pulchra*. *BMC Dev Biol.* 9:69.
- de Rosa R, et al. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399(6738):772–776.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279–D285.
- Fortunato SA, et al. 2014. Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. *Nature* 514(7524):620–623.
- Freyd G, Kim SK, Horvitz HR. 1990. Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene lin-11. *Nature* 344(6269):876–879.
- Fritzsch G, et al. 2008. PCR survey of *Xenoturbella bocki* Hox genes. *J Exp Zool B Mol Dev Evol.* 310(3):278–284.
- Frobis AC, Funch P. 2017. Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. *Nat Commun.* 8(1):9.
- Galliot B, de Vargas C, Miller D. 1999. Evolution of homeobox genes: q 50 Paired-like genes founded the Paired class. *Dev Genes Evol.* 209(3):186–197.
- García-Fernández J, Holland PW. 1994. Archetypal organization of the amphioxus Hox gene cluster. *Nature* 370(6490):563–566.
- Gavilán B, Perea-Atienza E, Martínez P. 2016. Xenacoelomorpha: a case of independent nervous system centralization? *Philos Trans R Soc Lond B Biol Sci.* 371(1685):20150039.
- Gehring WJ, Affolter M, Bürglin T. 1994. Homeodomain proteins. *Annu Rev Biochem.* 63:487–526.
- Gellon G, McGinnis W. 1998. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* 20(2):116–125.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–U130.
- Grau-Bové X, et al. 2017. Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife.* 6: pii: e26036.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hejnal A, Martindale MQ. 2009. Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutiloba longifissura*. *BMC Biol.* 7:65.
- Hejnal A, et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci.* 276(1677):4261–4270.
- Hench J, et al. 2015. The Homeobox genes of *Caenorhabditis elegans* and insights into their spatio-temporal expression dynamics during embryogenesis. *PLoS One* 10(5):e0126947.
- Herr W, et al. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.* 2(12A):1513–1516.
- Hobert O, Westphal H. 2000. Functions of LIM-homeobox genes. *Trends Genet.* 16(2):75–83.
- Holland LZ. 2015. The origin and evolution of chordate nervous systems. *Philos Trans R Soc Lond B Biol Sci.* 370(1684): 20150048.
- Holland LZ, et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* 18(7):1100–1111.
- Holland PW. 2013. Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol.* 2(1):31–45.
- Holland PW. 2015. Did homeobox gene duplications contribute to the Cambrian explosion? *Zool Lett.* 1:1.
- Holland PW, Booth HA, Bruford EA. 2007. Classification and nomenclature of all human homeobox genes. *BMC Biol.* 5:47.
- Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. 2017. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Philos Trans R Soc Lond B Biol Sci.* 372(1713): 20150480.
- Hyman LH. 1951. The invertebrates: Platyhelminthes and Rhynchocoela. The acoelomate bilateria. New York City (NY): McGraw-Hill.
- Jimenez-Guri E, Paps J, Garcia-Fernandez J, Salo E. 2006. Hox and ParaHox genes in Nemertodermatida, a basal bilaterian clade. *Int J Dev Biol.* 50(8):675–679.
- King N, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451(7180): 783–788.
- Kusserow A, et al. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433(7022):156–160.
- Larroux C, et al. 2007. The NK homeobox gene cluster predates the origin of Hox genes. *Curr Biol.* 17(8):706–710.
- Larroux C, et al. 2008. Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol.* 25(5):980–996.
- Loytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol Biol.* 1079:155–170.
- Matassi G, Imai JH, Di Gregorio A. 2015. Molecular phylogeny of four homeobox genes from the purple sea star *Pisaster ochraceus*. *Dev Genes Evol.* 225(6):359–365.
- McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ. 1984. A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* 308(5958):428–433.
- Mesika A, Ben-Dor S, Laviad EL, Futerman AH. 2007. A new functional motif in Hox domain-containing ceramide synthases: identification of a novel region flanking the Hox and TLC domains essential for activity. *J Biol Chem.* 282(37):27366–27373.
- Moreno E, Martínez P. 2010. Origin of bilaterian Hox patterning system. *eLS, Evolution & Diversity of Life.*



- Moreno E, Nadal M, Baguña J, Martínez P. 2009. Tracking the origins of the bilaterian Hox patterning system: insights from the acoel flatworm *Symsagittifera roscoffensis*. *Evol Dev*. 11(5):574–581.
- Paps J, Xu F, Zhang G, Holland PW. 2015. Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. *Genome Biol Evol*. 7(3):677–688.
- Philippe H, et al. 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* 470(7333):255–258.
- Ramachandra NB, Gates RD, Ladurner P, Jacobs DK, Hartenstein V. 2002. Embryonic development in the primitive bilaterian *Neochildia fusca*: normal morphogenesis and isolation of POU genes Brn-1 and Brn-3. *Dev Genes Evol*. 212(2):55–69.
- Rouse GW, Wilson NG, Carvajal JI, Vrijenhoek RC. 2016. New deep-sea species of Xenoturbella and the position of Xenacoelomorpha. *Nature* 530(7588):94–97.
- Ruiz-Trillo I, Riutort M, Littlewood DT, Herniou EA, Baguna J. 1999. Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science* 283(5409):1919–1923.
- Ryan AK, Rosenfeld MG. 1997. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev*. 11(10):1207–1225.
- Ryan JF, et al. 2006. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biol*. 7(7):R64.
- Ryan JF, et al. 2010. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* 1(1):9.
- Scott MP, Weiner AJ. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc Natl Acad Sci U S A*. 81(13):4115–4119.
- Seo HC, Curtiss J, Mlodzik M, Fjose A. 1999. Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. *Mech Dev*. 83(1–2):127–139.
- Srivastava M, et al. 2008. The Trichoplax genome and the nature of placozoans. *Nature* 454(7207):955–960.
- Srivastava M, Mazza-Curll KL, van Wolfswinkel JC, Reddien PW. 2014. Whole-body acoel regeneration is controlled by Wnt and Bmp-Admp signaling. *Curr Biol*. 24(10):1107–1113.
- Takatori N, et al. 2008. Comprehensive survey and classification of homeobox genes in the genome of amphioxus, *Branchiostoma floridae*. *Dev Genes Evol*. 218(11–12):579–590.
- Thomas-Chollier M, Martinez P. 2016. Origin of metazoan patterning systems and the role of ANTP-class homeobox genes. *eLS, Evolution & Diversity of Life*.
- Westblad E. 1949. Xenoturbella bocki n.g. n.sp., a peculiar, primitive turbellarian type. *Arkiv Zool*. 1:3–29.
- Zheng C, Chalfie M. 2016. Securing neuronal cell fate in *C. elegans*. *Curr Top Dev Biol*. 116:167–180.
- Zhong YF, Holland PW. 2011a. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. *BMC Evol Biol*. 11:169.
- Zhong YF, Holland PW. 2011b. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*. 13(6):567–568.

**Associate editor:** Dorothée Huchon